

Culturally Competent Assessment: More Than Nonbiased Tests

Russell J. Skiba, Ph.D.,^{1,4} Kimberly Knesting, Ph.D.,² and Lakeisha D. Bush³

The persistent and unresolved issue of minority disproportionality in special education provides a strong rationale for ensuring that assessment is culturally appropriate and sensitive. An extensive literature on test bias has failed to find evidence of bias sufficient to fully explain disproportional representation of minorities in special education. Yet minorities in this nation, most especially African Americans, have been exposed to a long history of unequal opportunity and oppression, and these inequities continue to be played out in educational settings. In the face of inadequate cultural and educational opportunity, unbiased tests provide an accurate estimate not only of individual capability, but also of the inhospitable conditions that depress that capability. A failure to take differential educational opportunity into account when considering minority test performance may lead to serious errors of test score interpretation. In the face of unequal educational opportunity, culturally competent assessment thus means more than manipulating the content of tests. Rather, it represents a comprehensive process that uses the results of assessment to identify and remedy educational conditions that systematically disadvantage students of color.

KEY WORDS: special education; disproportionate representation; assessment; evaluation; test bias.

The disproportionate representation of minorities in special education is among the most persistent and unresolved of problems in the field of special education. Despite more than thirty years of attention and debate on the topic, evidence

¹Associate Professor, Department of Counseling and Educational Psychology, Indiana University, Bloomington, IN.

²Assistant Professor, Department of Psychology, University of Wisconsin – Eau Claire, Eau Claire, WI.

³Doctoral Student, Department of Counseling and Educational Psychology, Indiana University, Bloomington, IN.

⁴Correspondence should be directed to Russell J. Skiba, Institute for Child Study, Rm. 0026, W. W. Wright Education Building, 201 N. Rose Ave., Bloomington, IN 47405-1006; e-mail: skiba@indiana.edu.

suggests that significant disproportionalities remain (Coutinho & Oswald, 2000; Dunn, 1968; National Research Council, 2000).

The debate about minority disproportionality in special education is often framed in terms of test bias: whether aptitude and achievement tests are constructed in such a way as to be inherently biased against certain groups, yielding inaccurate scores or inaccurate predictions. Yet the test itself is only one part of the assessment process; in order to be culturally competent, tests must also be administered and interpreted without bias. Indeed, culturally competent assessment might be defined as a process of assessment that does not contribute to the overrepresentation of minority students in special education.

In this paper, we explore the evidence concerning possible sources of bias in test construction, administration, and interpretation that may pose a threat to culturally competent assessment. Ultimately, we argue that culturally competent assessment is much more than ensuring that one's tests are unbiased. Rather, culturally competent assessment represents a commitment to data collection that does not contribute to overidentification, but rather assists in identifying and eliminating sources of bias throughout the educational process.

DISPROPORTIONATE MINORITY REPRESENTATION IN SPECIAL EDUCATION

The overrepresentation of minorities in special education has been an important and persistent topic almost since the inception of special education. Dunn (1968) first suggested that the over-representation of ethnic and language minority students in self-contained special education classrooms raised significant civil rights and educational concerns. Concerns about racial inequity were in fact central to litigation (e.g., *Mills v. Board of Education*, 1972) that led to the promulgation of the first special education legislation (P.L. 94-142, 1975).

Although there has been some disagreement about the meaning and extent of disproportionality (Oswald et al., 1999), concerns about overrepresentation continue. Analyses of Office of Civil Rights data have presented a relatively consistent picture of disproportionality (Reschly, 1997), especially in the categories of mild mental retardation and serious emotional disturbance. In response, the 1997 reauthorization of the Individuals with Disabilities Education Act (IDEA 97) stresses the importance of efforts to "prevent the intensification of problems connected with mislabeling and high dropout rates among minority children with disabilities."

ARE STANDARDIZED TESTS BIASED?

Perhaps due to the central role that standardized tests have played in important court tests concerning minority overrepresentation (*Larry P. v. Riles*, 1979; *PASE*

v. *Hannon*, 1980), a great deal of the research in this area has focused on the extent to which tests are biased against minority students. Brown, Reynolds and Whitaker (1999) define cultural test bias as the hypothesis that racial or ethnic differences in test scores reflect “no real differences in ability, but rather problems in the construction, design, administration, or interpretation of tests” (p. 209). Specific sources of psychometric bias could depress the scores of minority students relative to their expected “true score,” which in turn would yield the over-identification of minority students for special education service. A number of possible sources of test bias have been identified, including construct validity, content or item bias, inappropriate standardization samples, examiner and language bias, and differential predictive validity.

Construct Validity

It has been suggested that measures of intelligence lack construct validity, since IQ tests used with minorities are measuring not inherent aptitude, but rather the extent to which test-takers share white middle-class knowledge and values (Mercer, 1979). A standard methodology for testing differential construct validity involves comparing the factor structure of a test with different subgroups to determine if items have different meanings and relationships with other items for different subgroups. To date, there has been little evidence of a different factor structure for intelligence tests for black and white students, suggesting comparable construct validity across ethnic groups (Brown et al., 1999).

Sampling

Harrington (1975) has argued that sampling subpopulations proportionally will lead to tests that are biased against minority populations, since by definition minority populations are undersampled. Thus any test or task will favor the group which represents the largest proportion of the test construction sample. Animal analogue experiments (Harrington, 1984) appear to validate this hypothesis. Yet replications with human samples on intelligence and achievement tests have not replicated Harrington’s results (Hickman & Reynolds, 1987).

Predictive Validity

Consistently lower means of minorities have raised questions concerning the predictive validity of standardized tests for minority children. Yet mean group differences do not necessarily demonstrate lack of validity, if prediction is equally accurate across groups. Rather, predictive validity is better assessed using a regression approach, in which a test can be termed biased if it yields systematic overprediction or underprediction for a particular group (Brown et al., 1999). Regression

approaches to bias have not in general yielded evidence of differential predictive validity for minority and white students (Brown, et al., 1999; Cole, 1981; Weiss & Prifitera, 1995), although some evidence of bias in prediction has been reported in studies involving Mexican-American students (Suzuki & Valencia, 1997).

Item Bias

Lack of exposure of minority groups to information required by certain test items or subtests may result in items that are biased against group members. Consideration of intelligence test items proved to be an essential part of the case showing test bias in Larry P (Reschly, 1997). Yet statistical comparisons of African-American and white performance on a wide range of tests have generally failed to find a significant number of items that are biased against black students (Brown et al., 1999). Shepard (1987) argues, however, that analysis at the individual item level may be insufficient for exploring test bias: more sophisticated methodologies, such as item response theory, have yielded patterns of bias that explain a small but significant portion of the variance in black-white test score discrepancies.

Language and Examiner Bias

Language differences and the status of the examiner may also contribute to bias in testing. Taylor and Lee (1987) argue that, for those with nonstandard language experiences, tests which use language to assess psychological, social, or cognitive functioning often measure knowledge of Standard English rather than aptitude or achievement. The examiner may also represent a source of bias. The examiner's ethnic and social class prejudices may interfere with assessment practices (Oakland, 1980), but examiner effects may also be less explicit. In a meta-analysis of the effects of examiner familiarity on test performance, Fuchs and Fuchs (1986) reported that examiner unfamiliarity, defined in part as membership in a different group from the examinee, had a significant impact on standardized test performance. In particular, this effect was greater when the examinees were of low socioeconomic status.

While the majority of examiner bias research has focused on standardized measures of aptitude and achievement, some studies suggest that behavioral ratings are at equal or greater risk for the transmission of social and educational bias than other sources of assessment information. Both Mexican-American (Elliott & Argulewicz, 1983) and African American students (Kelly, Bullock, & Dykes, 1977) receive disproportionately high ratings on behavioral rating scales, and ratings appear to differ between ratings completed by white teachers and those completed by teachers of color. While these results may or may not indicate racial bias in the ratings process, they suggest that teacher ratings are at least as sensitive as other sources of assessment data to sources of examiner bias.

Summary of Test Bias Research

Evidence of a certain degree of item bias in standardized testing has led some to disagree with Jensen's (1980) categorical assertion that the possibility of bias in cognitive testing has been eliminated (Shepard, 1987). Others have argued that the failure to represent some minority populations in test bias research suggests that the "door on test bias research should not be closed" (Suzuki & Valencia, 1997). Yet in general, research has tended not to support the hypothesis that standardized tests of intelligence and achievement contain inherent sources of cultural bias (Brown et al., 1999; Cole, 1981; Jensen, 1980; Rosenbach & Mowder, 1981; Suzuki & Valencia, 1997). While the possibility of bias has not been conclusively ruled out, those sources of cultural bias that have been identified do not appear sufficient at present to explain lower minority test performance and disproportionate placement in special education.

In particular, the possibility of examiner or rater effects in assessment has not been ruled out. Degree of examiner familiarity appears to influence on the outcomes of standardized assessment, especially for students of lower socio-economic status. In particular, there is evidence of racial bias in social-emotional ratings. Given that the category of serious emotional disturbance is typically among those demonstrating the highest levels of disproportionality, such findings argue strongly for more research on the possible contribution of ratings bias to minority overrepresentation in special education. Finally, the strong potential for examiner effects in assessment argues that having a test that is culturally unbiased does not necessarily guarantee that the test will be used in a culturally competent manner.

THE BROADER CONTEXT OF MINORITY OVERREPRESENTATION

Often the issue of test bias and disproportionate minority representation in special education is framed as if there are only two possible outcomes to this debate: (a) if there are mean differences between ethnic groups on standardized measures of intelligence and achievement, the tests yielding those scores must contain some sources of bias toward ethnic minorities (Dent, 1987), or (b) if extensive research fails to find technical bias, then measured differences between ethnic groups represent "real" differences in aptitude that may in fact have some genetic basis (Herrnstein & Murray, 1994). Clearly, however, there are a number of alternate hypotheses that might explain the complex relationships between aptitude, achievement, educational opportunity, and outcome that might lead to depressed minority performance and special education over-representation. A 1982 report under the auspices of the National Academy of Sciences (Heller, Holtzmann, & Messick, 1982) implicated referral and prereferral practices as one possible source of disproportionate minority placement in special education. More recent reports (Coutinho & Oswald, 2000; National Research Council, 2002) have suggested

that sociodemographic, environmental, and educational factors may contribute to the depressed achievement of minority students, making them more susceptible to special education referral and placement.

Presuming that lower mean scores on standardized tests are a valid representation of minority performance does not of necessity lead to a conclusion of unequal ethnic aptitude. There is sufficient data to plausibly hypothesize that the black-white test score gap (and thus minority disproportionality in special education) is due neither to test bias nor to genetic inferiority, but rather to what Wilson (1998) terms the "structure of inequality," that is, numerous sources of cultural and educational disadvantage that fall disproportionately on certain subpopulations in our culture. Two sources of data support this perspective: (a) demographic data suggesting that minority disproportionality is not unique to special education, but reflects an overall pattern of unequal opportunity throughout the lifespan for members of disadvantaged groups, and (b) ubiquitous and often overlooked sources of differential opportunity in schools and classrooms that may well magnify the educational deficits that disadvantaged students bring with them at school entry.

The Extent of Disproportionality

Extensive disadvantage based on ethnic and socioeconomic status raises concerns about social equity in America that reach well beyond special education. Children of color are more than twice as likely as white children to live in single-parent households or below the poverty line (Demographic Factors in American Education, 1995). Only 5.6% of white children who have lived in poverty have experienced it for longer than five years, as compared with 39% of black children (Duncan, Brooks-Gunn, & Klebanov, 1994). Between 1985 and 1992, the average number of white infant deaths per 1,000 live births increased from 6.1 to 6.9, while for blacks the number increased from 12.1 to 16.8 (National Center for Health Statistics, 1996). African-Americans are more often exposed to factors associated with extreme poverty: inadequate high density housing, crime, noise, psychosocial stresses, and lack of maternal medicine and nutritional care (Polednak, 1996). The disproportionate risk extends even to environmental concerns such as exposure to dust, soot, carbon monoxide, ozone, sulfur, sulfur dioxide and lead, and emissions from hazardous waste dumps (Suro, 1993).

Extensive economic disadvantage influences not only the general quality of life for people of color, but also the preparedness and performance of many minority students in schools (Duncan et al., 1994; Entwisle & Alexander, 1993). The family background and parenting variables associated with poverty may, in and of themselves, be sufficient to explain a large proportion of the black white test score gap. In arguing that socioeconomic factors were insufficient to account for more than one-third of black-white test score discrepancies, Herrnstein and Murray (1995) calculated that "the average environment of blacks would have to be at the 6th percentile of the distribution of environments among whites . . . for racial differences

to be entirely environmental” (p.299). Using a regression model to control for shared environmental-genetic variance, however, Phillips, Brooks-Gunn, Duncan, Klebanov, and Crane (1998) were able to show that a combination of mother’s educational attainment, family income, parenting practices, and neighborhood effects were able to account for a far larger proportion of variance in the test score gap than had previously been assumed. Together these variables suggest that the average black child’s “environment” may well rank at the 8th percentile of the white distribution.

EDUCATIONAL INEQUITY: MAGNIFYING SOCIOECONOMIC DISADVANTAGE

Unfortunately, there is also evidence suggesting that the gaps in verbal and mathematics ability between white and black students widen after school entrance. A meta-analysis of extant longitudinal and cross-sectional surveys of student achievement through all grade levels concluded that black students who start elementary school on a par with average white test scores will learn less than the average white student by twelfth grade (Phillips, Crouse & Ralph, 1998). Further the report concluded that, at the end of high school, 44% of the black-white gap in mathematics scores and 57% of the gap in reading scores cannot be attributed to differences between black and white students at school entry.

Such results do not automatically imply that school practices cause the widening of the racial gap. Yet there is abundant evidence of differential educational opportunity for black and white students throughout the educational process. A number of sources of structural inequality in schools have been documented, in the areas of physical facilities, curriculum, teacher expectations, school discipline, tracking, and quality of instruction.

Physical Facilities and Resource Inequality

Facilities in urban, lower socioeconomic catchment areas that are disproportionately minority in makeup appear to suffer from a number of structural disadvantages, including older, inadequate buildings, and out-of-date and insufficient supplies, curriculum, and equipment (Kozol, 1991; Oakes, 1990). Inadequate physical conditions, ranging from hot and stuffy classrooms to safety and health concerns, in turn reduce the effectiveness of the school environment for stimulating teaching and learning (Kozol, 1991).

Discrepancies in the quality of physical facilities and available educational resources in general may well be linked to a system of public school finance that favors high income districts at the expense of lower income urban districts (Dunn, 1999; Kozol, 1991). A wave of fiscal equity legislation has challenged school funding formulas that have a discriminatory impact on minority students (Dunn, 1999; Rebell, 1999; Singer, 1999), with some success.

Curriculum

School curriculum drives instruction and provides important messages to students about the status and history of the subgroups that comprise American culture. Yet available evidence suggests that commonly used curricula often act to reinforce race, gender, and economic stereotypes. Curricula and instruction in working class schools tends to be more procedural, emphasizing rote learning over application or problem solving, and presenting fewer images of diversity than curriculum in middle-class or affluent schools (Anyon, 1981). Surveys of standard curriculum for grades 1 to 8 have revealed that whites dominate the story line and history presented in most textbooks, and that the roles described for minorities and women are consistently more circumscribed than those of white males (Sleeter & Grant, 1991).

Teacher Expectations

Field-based studies have demonstrated that teacher expectations and ratings of student achievement in the early elementary years are predictive of achievement at the middle and high school level (Entwisle & Alexander, 1993). Lower expectations for poor African-American children were documented thirty years ago (Leacock 1969). Unfortunately, differential expectations for students of different racial or social classes continue to be documented (Murray, 1996) and those beliefs appear to be correlated with less effective instructional practice (Solomon, Battistich, & Hom, 1996).

School Discipline

Racial and socioeconomic disproportionality in the administration of school discipline has been widely documented. Black students are suspended at a rate between two and three times that of white students (Children's Defense Fund, 1975; McFadden, Marsh, Price & Hwang, 1992; Skiba, Peterson & Williams, 1997) and are significantly more likely to receive corporal punishment (Shaw & Braden, 1990). Ethnic biases in the administration of suspension appear to be independent of socioeconomic status and do not appear to be based solely on behavioral or attitudinal differences between black and white students (Wu, Pink, Crain, & Moles, 1982). If anything, black students appear to be punished more harshly for comparable offenses (McFadden et al., 1992), or referred for discipline for less objective offenses (Skiba, Michael, Nardo, & Peterson, in press).

Tracking and Within-Class Grouping

In an extensive study of mathematics and science education, Oakes (1990) reported "clear and consistent patterns of unequal opportunities to learn mathematics

and science” (p. 3) between students from different ethnic and socioeconomic groups. Much of this difference appears to be due to the effects of academic tracking: black students are disproportionately represented in lower tracks, which are often handicapped by lower quality curricula, equipment, and instruction. While recent research has shown that there may be more mobility among high school tracks than previously believed, black students are still disproportionately represented in lower tracks and are more likely to drop basic requirements in English and Mathematics (Hallinan, 1996). Welner and Oakes (1996) suggest that an end to racial disparities in tracking may require increased use of the type of litigation that has already been successful in curtailing such practices in several school districts.

Instructional Quality

Among the most significant discrepancies between affluent, predominantly white schools and low SES urban, predominantly black schools may be the quality of instruction. Less than adequate facilities, restrictive curriculum, and low morale often combine with poor student readiness to make teacher retention in low income urban areas highly problematic (Kozol, 1991); as a result, students in lower SES schools have less contact with well qualified teachers (Oakes, 1990). Greenwood, Delquadri, and Hall (1984) reported that the instructional strategies of teachers in inner-city schools were significantly less effective in promoting academic responding for low SES children.

Indirect Effects

Given the complexity of these factors, the relationship between disadvantage and educational outcomes may be mediated by any of a number of variables. Ogbu (1986) argues that the legacy of the “American caste system” has left many African-Americans with a deep distrust of American institutions, including education. African-American students may develop the attitude that school success is antithetical to their racial identity, choosing to fail to avoid “acting white.” Resistance theory (Skutnabb-Kangas & Cummins, 1988) suggests that minority students who believe that the educational system is failing to meet their needs may well act out these feelings through non-compliance, lack of attention, and poor academic performance.

Summary

In rejecting the argument that a background of bias, “uniform and ubiquitous,” can explain black-white test score discrepancies, Herrnstein and Murray (1994) assert that “...everyday experience suggests that the environment confronting blacks in different sectors of American life is not uniformly hostile” (p. 286).

Yet while the level of hostility toward ethnic minorities may have diminished in our culture, we have by no means attained the ideal of equal opportunity for all individuals in our culture, or equal educational opportunity in our schools. Beyond a failure to remediate disadvantage, public education appears to contain structural inequities that may magnify disadvantage, guaranteeing that children who come to school with differences or deficits will continue to fall further behind the longer they are in school.

These inequities have a demonstrable effect on the educational opportunity and school achievement of low SES and African-American children. In a multi-year observational study, Greenwood et al. (1994) reported that inferior instruction in low SES schools resulted in students in those schools receiving an equivalent of 57 weeks less academic engagement than students in high-SES schools by the sixth grade; as a result, an achievement gap equal to 0.3 of a grade level at school entry grew to a gap of 3.5 grade levels by grade 6. These data make a strong case that low-SES, African-American, and Hispanic students are at greater risk for poor quality educational experiences which may in turn significantly undermine their academic achievement. Successful legal challenges to discriminatory school practices in the areas of tracking (Welner & Oakes, 1996) and school funding (Rebell, 1999) suggest that the courts have decided that current educational practices in some areas *are* uniformly hostile toward minority youth.

WHAT DO TESTS REALLY MEASURE?

Intelligence and achievement tests appear to be as accurate in predicting the present and future performance of African-American students as they are in predicting the performance of white students (Brown et al., 1999; Heller et al., 1982). Depressed minority performance on standardized tests can be expected to be accurate in their predictions that students of certain backgrounds will on average fare more poorly in both school and society. Yet while depressed test scores are a predictor of future performance, they are also a *product* of cultural and educational disadvantage. Tests that are unbiased provide an accurate estimate of current individual aptitude; yet they also provide an unbiased and accurate record of the effects of unequal educational opportunity. A standardized test score may be a relatively reliable predictor of a disadvantaged student's current capabilities. But it also accurately summarizes the tragic history of limited cultural and educational opportunity that led to those depressed capacities.

The likely impact of cultural and educational conditions on minority academic performance has an extremely important implication for the interpretation of test scores. If *cultural and educational contributions to depressed minority scores are overlooked, those scores will become biased estimators of individual potential* by misattributing the effects of inadequate educational opportunity to a lack of individual aptitude or ability. Research has increasingly shown the critical contribution

of a child's opportunity to learn to academic achievement (Greenwood et al., 1984). Accurate interpretation of the test scores of disadvantaged students must thus take into account the negative effect of inadequate educational opportunity on those scores. Gould (1995) has argued that the outrage accompanying the publication of *The Bell Curve* was to some extent a reaction to the authors' overemphasis on "technical" test bias and genetic explanations, to the exclusion of pervasive cultural biases. Indeed, there seems to be something inherently unjust in treating test scores that reflect a persistent inequity in both school and culture as an accurate assessment of "individual differences."

Neglecting the contribution of educational opportunity when interpreting minority test scores may in fact introduce statistical error, regardless of whether a specific test can be shown to be "unbiased." Failure to consider an important source of variance represents what structural modeling theorists term "specification error" (Kenny, 1979); that is, by neglecting to include important variables in a model, that model may become inaccurate. If the test results of children whose educational opportunities have been limited are treated solely as an assessment of individual aptitude, variance due to educational disadvantage may be mistakenly attributed to individual capability. Failure to take into account issues of limited educational opportunity for disadvantaged students thus increases measurement error. As measurement error due to mismeasured variables increases, the reliability of the data decreases. In short, underestimating the effects of unequal opportunity on a minority child's test results is a serious error of interpretation that probably reduces the accuracy of predictions drawn from that data.

Unfortunately, concerns raised thus far about the interpretation of minority test scores may only be magnified by questions regarding the efficacy of special education. Reschly (1997) notes that disproportionate placement of minorities in remedial programs is a problem only if that program is not viewed as beneficial or remedial. Yet evidence such as poor graduation rates (Edgar, 1987) suggests that special education service has not been highly successful in overcoming educational disadvantage. It is probably revealing that there have been no private or class action suits in the courts seeking to prove reverse discrimination as a result of greater minority representation in special education in general, or SED programs in particular.

It has been argued that certain structures of American schools, such as tracking, serve to replicate and perpetuate ethnic and class distinctions, through a process termed *cultural reproduction* (Oakes, 1982). The use of standardized tests for determining special education eligibility for minority students may exemplify this process. Historical bias and current inequities continue to place minority students at an educational disadvantage. The results of unequal educational opportunity are then "validly" reflected in depressed scores for African-American and Hispanic students. These scores are in turn used to place minority students in special programs. To the extent that such programs fail to remediate educational disadvantage, testing has thus made an important contribution to the cultural reproduction

of social inequity. When minority assessment results are used uncritically by those who administer and interpret tests, those professionals may become unwitting participants in a continuing cycle of disadvantage.

DEVELOPING CULTURALLY COMPETENT ASSESSMENT

On one level, it is not surprising that an intensive search for test bias has failed to find test anomalies sufficient to account for ethnic disproportionality in test performance. As long as cultural and educational inequities systematically disadvantage entire classes of individuals, valid tests will accurately reflect the outcomes of those biases. Indeed, at this point in history, tests that *failed* to demonstrate some form of disadvantage for victims of racial or socioeconomic bias might be said to lack concurrent and predictive validity with respect to the wider culture.

A broader approach to the problem of test bias and minority overrepresentation has important implications for the definition of culturally competent assessment. Assuming that test bias constitutes the primary cause of disproportionality has led to a concern with the technology of testing, attempting to design culture-fair tests (Mercer, 1979). If, however, depressed minority performance on standardized measures is due not to faulty tests or procedures, but rather to the pervasive influence of differential educational opportunity, then no amount of adjustment of the assessment process will address the problem. Authentic or performance-based assessment, for example, has recently been seen as a viable alternative for reducing test bias. Yet accumulating evidence suggests that alternative assessment formats do not result in a significant reduction of the black-white test score discrepancy as compared with more traditional formats (Braden, 1999).

Culturally competent assessment, then, means a good deal more than simply adjusting test items. It is rather a comprehensive process that uses the results of assessment to identify and remedy educational conditions that systematically disadvantage certain groups of students. If cultural and educational bias play a role in ethnic disproportionality, then culturally competent assessment uses both individual and system-wide data, not to perpetuate social inequality, but rather to monitor progress toward the goal of educational equity. Some possible components of such a process are outlined below.

Ensure the Cultural Competence of the Examiner

Although test bias research has not in general found evidence of cultural bias in tests themselves, there is suggestive evidence of examiner effects that may lead to bias, especially in behavior ratings. Such findings strongly support the need for adequate training of those engaging in either traditional or behavioral assessment in culturally competent practices (Castillo, Quintana, & Zamarripa, 1999). Examples of critical skills necessary for the culturally competent administration

and interpretation of assessment measures include the ability to accurately test one's hypotheses about children, to avoid stereotyping, and to draw upon a broad understanding of diverse cultures (Sue, 1998).

Assess Disproportionate Representation

Data on the extent of minority over-representation in educational strategies or programs may provide important baseline information against which to judge system change efforts. The 1997 reauthorization of the Individuals with Disabilities Act requires states to collect data on the extent of minority overrepresentation in special education. At the local level, these data can act as a stimulus for administrators, school boards, and the public to begin a public dialogue concerning educational equity, and to monitor progress toward educational equity. There are a number of methods for quantifying discrepancies between white and minority enrollments in special education programs; readers interested in the benefits and drawback of different methods of calculating ethnic representation are encouraged to consult Reschly (1997).

Discrepancies between African-American and white scores on standardized tests declined by about a third of a standard deviation between 1976 and 1990; these changes may be due in part to shifts in public policy (Grissmer, Flanagan, & Williamson, 1998). If so, group averages in intelligence and achievement may, over time, provide an indicator of the extent to which efforts at reducing educational inequity have been successful in reducing the measured gap between black and white outcomes.

Assess Educational Opportunity

Districts interested in increasing cultural competence may wish to assess specific sources of inequity or bias that limit the educational opportunity of disadvantaged students. Williams (1989) has recommended district level data collection on minority overrepresentation in school discipline procedures as a method of monitoring progress toward eliminating those discrepancies. That same approach might be used to evaluate any of a number of indicators of educational opportunity. To what extent are children of color disproportionately placed in lower-track academic programs? Do the curricula currently in use at the district level adequately represent the history, contributions, and potential of the diverse cultures in this nation? Is the quality of resources and school structures consistent across socioeconomic lines, or do school facilities in lower SES communities reflect (and thus perpetuate) the poverty of the surrounding community? To the extent that these measures indicate diminished educational opportunity for a given group of students, culturally competent assessment will weigh those factors in the interpretation of test scores for that group.

Quality of instruction may be the single most important indicator of educational opportunity (Coleman, 1966). In the last 15 years, a technology for assessing the quality of classroom instruction has become available (Roberts, 1995), through both ratings (Ysseldyke & Christensen, 1987), and direct observation (Greenwood et al., 1984). If minority test scores are depressed, data on instructional quality can provide important information on the extent to which lower scores are due to decreased instructional opportunity.

Assess the Quality of Remedial and Compensatory Programs

Overrepresentation of any group in a given program is only a problem if the quality of that program is in doubt. If tests that reflect inadequate educational opportunity are used to justify placement in programs that do not address that disadvantage, both the test and the placement contribute to the perpetuation of educational inequality. A number of outcome indicators might be used to assess the adequacy of remedial programming. What proportion of ethnic minority students are decertified from special education, and how quickly? Are students of color served primarily in less or more restrictive environments than special education students in general? Do special and remedial education programs appear to have a positive, or an exacerbating, influence on dropout or graduation rates of minority students?

Assess Academic Experience and Motivation

Models of minority underachievement suggest that consistent cultural and educational disadvantage may result in minority students distrusting the educational process, and doubting the value of educational success (Ogbu, 1986). Thus, there may be a lack of motivation for, or even resistance to (Skutnabb-Kangas & Cummins, 1988), academic success. In order to assess the indirect effects of educational inequity, it may be necessary to assess minority student attitudes and experiences with respect to education. This type of data might well be used to assess the extent to which motivational variables serve to depress minority student performance.

Implement Low Inference Assessment Strategies

Ultimately, disproportionality of minorities in categories is dependent on the use of a classification-based model of remediation. There is increasing movement toward the use of direct measurement approaches that reduce the level of inference involved in assessment (Shapiro, 2000). Direct assessment approaches, such as curriculum-based measurement (Shinn, 1995) and functional assessment (Lane,

Umbreit, & Beebe-Frankenberger, 1999) are intended less for eligibility determination and placement than for intervention design and evaluation. Assessment information used to plan or evaluate remediation, without reference to placements, labels, or hypothetical underlying processes, may short circuit the issue of overrepresentation, since the issue of classification and categories does not arise. Such information provides a straightforward estimate of current educational needs and, more importantly, generates interventions that may increase educational opportunity.

In its recently released report on minority placement in special and gifted education, the National Research Council (2002) suggests that school psychologists can be a valuable resource in implementing these and other direct assessment approaches, including classroom observation and multiple gating approaches for early screening.

CONCLUSIONS

One of the problems stemming from the historical overemphasis on test bias is a complacency that appears to arise when statistical bias cannot be demonstrated. In an otherwise cogent review of test bias, Rosenbach and Mowder (1981) concluded that the absence of bias in psychoeducational measures means that “the problem is not one of psychological origin and is not likely to yield to psychological or empirical attacks. The cultural bias of schooling is philosophically based, and, ultimately, can be resolved only through socio-political actions. Unfortunately, this places educational professionals such as school psychologists in an unenviable position, for there is little they can do in their professional roles other than to deal with the issue at a superficial level” (p. 453).

We must respectfully disagree. The absence of technical bias in intelligence and achievement tests in no way absolves those who administer and make decisions based on those tests from socially responsible decision-making. Standardized tests appear to be relatively accurate in assessing both individual aptitude in an unbiased fashion. Yet they also accurately describe the broad discrepancy in the adequacy of educational opportunity between majority and minority populations in this country. Test scores provide an index of both individual capability and any ways in which the development of that capability has been limited. If test results are treated solely as a representation of inherent individual contribution to aptitude, ignoring cultural and educational factors that depress minority performance, test scores may yield inaccurate interpretations, even if the tests themselves can be shown to be unbiased. Using these results to place students in programs that fail to remediate educational disadvantage makes testing a mechanism for the replication and reinforcement of social inequality. Culturally competent assessment goes well beyond modifying tests. Cultural competence in assessment is based on awareness of the social and historical forces that continue to depress the academic performance of minorities,

and reaches fruition in a systems change effort wherein the process and results of assessment are used to break rather than perpetuate the cycle of social inequity.

REFERENCES

- Anyon, J. (1981). Social class and school knowledge. *Curriculum Inquiry*, 11, 3–42.
- Artiles, A. J., & Trent, S. C. (1994). Overrepresentation of minority students in special education: A continuing debate. *Journal of Special Education*, 27, 410–437.
- Braden, J. P. (1999). Performance assessment and diversity. *School Psychology Quarterly*, 14, 304–326.
- Brown, R. T., Reynolds, C. R., & Whitaker, J. S. (1999). Bias in mental testing since *Bias in Mental Testing*. *School Psychology Quarterly*, 14, 208–238.
- Castillo, E. M., Quintana, S. M., & Zamarripa, M. X. (1999). Cultural and linguistic issues. In E. S. Shapiro & T. R. Kratochwill (Eds.), *Conducting school-based assessments of child and adolescent behavior* (pp. 274–308). New York: Guilford.
- Children's Defense Fund. (1975). *School suspensions: Are they helping children?* Cambridge, MA: Author.
- Cole, N. S. (1981). Bias in testing. *American Psychologist*, 36, 1067–1077.
- Coleman, J. S. (1966). *Equality of educational opportunity*. Washington, DC: U. S. Office of Education.
- Coutinho, M. J., & Oswald, D. P. (2000). Disproportionate representation in special education: A synthesis and recommendations. *Journal of Child and Family Studies*, 9, 135–156.
- Demographic factors in American education*. (1995). Arlington, VA: Educational Research Service. (ERIC Document Reproduction Service No. ED 379 773).
- Dent, H. E. (1987). The San Francisco Public Schools experience with alternatives to IQ testing: A model for non-biased assessment. *The Negro Educational Review*, 38 (2–3), 146–162.
- Duncan, G. J., Brooks-Gunn, J., & Klebanov, P. K. (1994). Economic deprivation and early childhood development. *Child Development*, 65, 296–318.
- Dunn, C. (1999). Challenging racial discrimination in the public schools of New York State. *Equity and Excellence in Education*, 32, 19–24.
- Dunn, L. M. (1968). Special education for the mildly retarded—is much of it justifiable? *Exceptional Children*, 23, 5–21.
- Edgar, E. (1987). Secondary programs in special education: Are many of them justifiable? *Exceptional Children*, 53, 555–561.
- Elliot, S. N., & Argulewicz, E. N. (1983). The influence of student ethnicity on teachers' behavior ratings of normal and learning disabled children. *Hispanic Journal of Behavioral Sciences*, 5, 337–345.
- Entwisle, D. R., & Alexander, K. L. (1993). Entry into school: The beginning school transition and educational stratification in the United States. *Annual Review of Sociology*, 19, 401–423.
- Fuchs, D., & Fuchs, L. S. (1986). Test procedure bias: A meta-analysis of examiner familiarity effects. *Review of Educational Research*, 56, 243–262.
- Gould, S. J. (1995). Mismeasure by any measure. In R. Jacoby, & N. Glauber (Eds.), *The Bell Curve debate: History, documents, opinions* (pp. 3–13). New York: Times Books.
- Greenwood, C. R., Delquadri, J. C., & Hall, R. V. (1984). Opportunity to respond and student academic performance. In W. L. Heward, T. E. Heron, D. S. Hill, & J. Trap-Porter (Eds.), *Focus on behavior analysis in education*. Columbus, OH: Charles E. Merrill.
- Grissmer, D., Flanagan, A., & Williamson, S. (1998). Why did the black-white score gap narrow in the 1970's and 1980's? In C. Jencks & M. Phillips (Eds.), *The black-white test score gap*. (pp. 182–226). Washington, D.C.: Brookings Institution Press.
- Hallinan, M. T. (1996). Track mobility in secondary school. *Social Forces*, 74, 983–1002.
- Harrington, G. M. (1975). Intelligence tests may favour the majority groups in a population. *Nature*, 258, 708–709.
- Harrington, G. M. (1984). An experimental model of bias in mental testing. In C.R. Reynolds & R. T. Brown (Eds.), *Perspectives on bias in mental testing* (pp.101–138). New York: Plenum.
- Heller, K. A., Holtzmann, W. H., & Messick, S. (Eds.). (1982). *Placing children in special education: A strategy for equity*. Washington, D. C.: National Academy Press.

- Herrnstein, R. J., & Murray, C. (1994) *The Bell Curve: Intelligence and class structure in American life*. New York: Free Press.
- Hickman, J. A., & Reynolds, C. R. (1987). Are race differences in mental test scores an artifact of psychometric methods?: A test of Harrington's experimental model. *Journal of Special Education*, 20, 409–430.
- Jensen, A. R. (1980). *Bias in mental testing*. New York: Free Press.
- Kelly, T. J., Bullock, L. M., & Dykes, M. K. (1977). Behavioral disorders: Teachers' perceptions. *Exceptional Children*, 43, 316–318.
- Kozol, J. (1991). *Savage inequalities: Children in America's schools*. New York: Harper.
- Lane, K. L., Umbreit, J., & Beebe-Frankenberger, M.E. (1999). Functional assessment research on students with or at risk for EBD: 1990 to the present. *Journal of Positive Behavior Interventions*, 1, 101–111.
- Larry P. v. Riles, 495 F.Supp. 926 (N.D. Cal. 1979).
- Leacock, E. B. (1969). *Teaching and learning in city schools: A comparative study*. New York: Basic Books.
- McFadden, C. A., Marsh, E. G., Price, J. Z., & Hwang, Y. (1992). A study of race and gender bias in the punishment of school children. *Educaton and Teatment of Children*, 5, 140–146.
- Mercer, J. R. (1979). In defense of racially and culturally nondiscriminatory assessment. *School Psychology Digest*, 8, 89–115.
- Mills v. Board of Education of the District of Columbia, 384 F. Supp. 866 (1972).
- Murray, C. B. (1996). Estimating achievement performance: A confirmation bias. *Journal of Black Psychology*, 22, 67–85.
- National Center for Health Statistics. (1996). *Vital statistics of the United States. 1992: Vol II. Mortality, Part B*. Washington, D.C.: Public Health Service.
- National Research Council (2002). *Minority students in special and gifted education*. Committee on Minority Representation in Special Education. Washington, D.C.: National Academy Press.
- Oakes, J. (1982). The reproduction of inequity: The content of secondary school tracking. *The Urban Review*, 14, 107–120.
- Oakes, J. (1990). *Multiplying inequalities: The effects of race, social class, and tracking on opportunities to learn mathematics and science*. Washington, D.C.: National Science Foundation.
- Oakland, T. (1980). Nonbiased assessment of minority group children. *Exceptional Education Quarterly*, 1, 31–46.
- Ogbu, J. U. (1986). The consequences of the American caste system. In U. Neisser (Ed.), *The school achievement of minority children: New perspectives*. Hillsdale, NJ: Lawrence Erlbaum.
- Oswald, D. P., Coutinho, M. J., Best, A. M., & Singh, N. N. (1999). Ethnic representation in special education: The influence of school-related economic and demographic variables. *The Journal of Special Education*, 32, 194–206.
- PASE (Parents in Action on Special Education) v. Joseph P. Hannon, U.S. District Court, Norther District of Illinois, Eastern Division, No. 74 (3586), July, 1980.
- Phillips, M., Brooks-Gunn, J., Duncan, G. J., Klebanov, P., & Crane, J. (1998). Family background, parenting practices, and the black-white test score gap. In C. Jencks & M. Phillips (Eds.), *The black-white test score gap*. (pp. 103–145). Washington, D.C.: Brookings Institution Press.
- Phillips, M., Crouse, J., & Ralph, J. (1998). Does the black-white test score gap widen after children enter school? In C. Jencks & M. Phillips (Eds.), *The black-white test score gap*. (pp. 229–272). Washington, D.C.: Brookings Institution Press.
- Polednak, A. P. (1996). Trends in U.S. urban black infant mortality by degree of residential segregation. *American Journal of Public Health*, 86, 723–726.
- Rebell, M. A. (1999). Fiscal equity litigation and the democratic imperative. *Equity and Excellence in Education*, 32, 5–18.
- Reschly, D. J. (1997). *Disproportionate minority representation in general and special education programs: Patterns, issues, and alternatives*. Des Moines, IA: Mountain Plains Regional Resource Center.
- Roberts, M. L. (1995). Best practices in assessing environmental factors that impact student performance. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology: III* (pp. 679–688). Washington, D. C.: National Association of School Psychologists.

- Rosenbach, J. H., & Mowder, B. A. (1981). Test bias: The other side of the coin. *Psychology in the Schools*, 18, 450–454.
- Shapiro, E. S. (2000). School psychology from an instructional perspective: Solving big, not little problems. *School Psychology Review*, 29, 560–572.
- Shaw, R. S., & Braden, P. J. (1990). Race and gender bias in the administration of corporal punishment. *School Psychology Review*, 19, 378–383.
- Shepard, L. A. (1987). The case for bias in tests of achievement and scholastic aptitude. In S. Modgil & C. Modgil (Eds.), *Arthur Jensen: Consensus and controversy* (pp. 177–190). New York: Falmer Press.
- Shinn, M. R. (1995). Best practices in curriculum-based measurement and its use in a problem-solving model. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology: III* (pp. 547–567). Washington, D. C.: National Association of School Psychologists.
- Singer, A. (1999). American apartheid: Race and the politics of school finance on Long Island, NY. *Equity and Excellence in Education*, 32, 25–35.
- Skutnabb-Kangas, T., & Cummins, J. (1988). *Minority education: From shame to struggle*. Clevedon, England: Multilingual Matters.
- Skiba, R. J., Michael, R., Nardo, A., & Peterson, R. (in press). The color of discipline: Sources of racial and gender disproportionality in school punishment. *The Urban Review*.
- Skiba, R. J., Peterson, R. L., & Williams, T. (1997). Office referrals and suspension: Disciplinary intervention in middle schools. *Education and Treatment of Children*, 20, 295–315.
- Sleeter, C. E., & Grant, C. A. (1991). Race, class, gender and disability in current textbooks. In M. W. Apple & Y. L. K. Christian-Smith (Eds.), *The politics of the textbook*. New York: Routledge & Chapman-Hall.
- Solomon, D., Battistich, V., & Hom, A. (1996). Teacher beliefs and practices in schools serving communities that differ in socioeconomic level. *Journal of Experimental Education*, 64, 327–347.
- Sue, S. (1998). In search of cultural competence in psychotherapy and counseling. *American Psychologist*, 53, 440–448.
- Suro, R. (1993, January 11). Pollution-weary minorities try civil rights tack. *The New York Times National*, pp. A1, A12.
- Taylor, O. L., & Lee, D. L. (1987). Standardized tests and African-American children: Communication and language issues. *The Negro Educational Review*, 38, 67–80.
- Weiss, L., & Prifitera, A. (1995). An evaluation of differential prediction of WIAT achievement scores from WISC-III FSIQ across ethnic and gender groups. *Journal of School Psychology*, 33, 297–304.
- Welner, K. G., & Oakes, J. (1996). (L)Ability grouping: The new susceptibility of school tracking systems to legal challenges. *Harvard Educational Review*, 66, 451–472.
- Williams, J. (1989). Reducing the disproportionately high frequency of disciplinary actions against minority students: An assessment-based policy approach. *Equity and Excellence*, 24(2), 31–37.
- Wilson, W. J. (1998). The role of the environment in the black-white test score gap. In C. Jencks & M. Phillips (Eds.), *The black-white test score gap*. (pp. 501–510). Washington, D.C.: Brookings Institution Press.
- Wu, S. C., Pink, W. T., Crain, R. L., & Moles, O. (1982). Student suspension: A critical reappraisal. *The Urban Review*, 14, 245–303.
- Ysseldyke, J. E., & Christenson, S. L. (1987). Evaluating students' instructional environments. *Remedial and Special Education*, 8 (3), 17–24.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.